

A Method and System for Enhanced User Experience of Audio

REFERENCE TO RELATED APPLICATIONS

- 5 **[0001]** This application claims priority from co-pending U.S. Provisional Application Serial No. 60/263,415 entitled 'A Method and System for Enhanced Streaming of Audio to Telephone', filed January 22, 2001.

FIELD OF THE INVENTION

- 10 **[0002]** The present invention relates to a system and method for audio buffering and navigation.

BACKGROUND OF INVENTION

- 15 **[0003]** Users' experience of media content such as films, music and lectures, has been enhanced through advances in audio technology. Technological advances have improved audio transmission quality and provided tools to the user that enhance the user's experience.

- 20 **[0004]** One of the key technological enhancements in audio transmission is 'streaming'. Streaming can be defined as a technique for transferal of data while continually processing the data. For example, a computer can take advantage of streaming technology to render audio during a download. Streaming is thus useful for users who wish to listen to extensive sections of audio without having to wait for transfer of the entire data. Rather than downloading the file in its entirety before playing, users can start listening to the audio before the transmission is completed

- 25 **[0005]** When streaming audio over a network, traffic problems are often encountered and the delivery of a steady stream of packets to the client is not always possible. During these periods of congestion, the user may experience disruption in service. For example, should network congestion occurs when using a REAL AUDIO ® player, the player will cease rendering audio and display the words 'buffering'
- 30 accompanied by flashing red and green lights (see www.real.com). This buffering occurs when the rate at which packets arrive from the server to the client is slower than the rate from client to user. To limit the occurrence of interruption, streaming clients

employ a buffer, typically embodied in a segment of memory, of most recently received audio packets.

[0006] Typically the network may slow down or speed up for periods of time, corresponding to low or high bandwidth availability. During times of high bandwidth, the buffer is filled; when transmission degrades, packets drawn from the buffer ensure a smooth rendering from the client to the user. The buffer operates in a First In First Out (FIFO) manner, whereby only the most recent bytes are stored. This is accomplished by constantly playing from the buffer while simultaneously inserting new bytes.

[0007] As the frequency of network problems increase, the contents of the buffer gradually decrease. When the buffer has been depleted, playing may cease.

[0008] Existing technology for navigation within audio samples provides users with basic navigational functions such as 'fast forward', 'rewind' and 'pause'. In addition to these basic functions, current technologies are capable of more advanced functions such as 'time compression', which enables the speeding up or slowing down of audio content without changing the pitch or audio quality. For instance, Creative Technology Ltd's EAX Technology® offers advanced audio functionality and performance for digital audio. EAX Technology® includes time-scaling adjustments in order to change the speed of the audio to suit the user's individual preferences.

[0009] In addition to time-scaling, InterVideo, Inc has produced a DVD player that possesses a "time stretching" feature. "Time stretching" allows a user to play a three hour DVD film in two hours, by adjusting the playback speed from half-speed to double-speed while still maintaining the original audio quality of the film.

[0010] Another advanced audio product is embodied in a DIR911 audio processor of Eventide Inc., which features an Intelli-Clear ® speed control. The Eventide audio processor DIR911 provides the option to rewind or fast-forward an answer phone message in small segments of 1.25 seconds at a time.

[0011] "SpeechSkimmer" is a user interface for skimming speech recordings developed by the Speech Research Group at the MIT Media Laboratory. "SpeechSkimmer" enables a user to hear recorded sounds quickly at several levels of detail by using simple speech processing techniques. *"A continuum of time compression and skimming techniques have been designed, allowing a user to efficiently skim a speech recording to find portions of interest, then listen to it time-compressed to allow*

quick browsing of the recorded information, and then slowing down further to listen to detailed information.” (Barry Arons ‘SpeechSkimmer: Interactively Skimming Recorded Speech. Proceedings of UIST 1993: ACM Symposium on User Interface Software and Technology. ACM Press. Nov 3-5 1993. Atlanta.)

5

SUMMARY OF THE INVENTION

[0012] This invention seeks to provide a system and method for audio buffering and navigation.

10 **[0013]** There is thus provided in accordance a system for providing enhanced quality audio streaming. The system includes an audio streaming server providing an audio stream, a client including a buffer storing at least portions of the audio stream received from the audio streaming server, a buffer status sensor operative to monitor the contents of the buffer and a client audio output enhancer which operates in response to an output from the buffer status sensor for providing a modified audio stream output.

15 **[0014]** There is also provided in accordance with a preferred embodiment of the present invention a method for providing enhanced quality audio streaming. The method includes the steps of: providing an audio stream to a client, storing in a buffer at least portions of the audio stream, monitoring contents of the buffer and providing a modified audio stream in response to an output from the monitoring.

20 **[0015]** Further in accordance with a preferred embodiment of the present invention the client audio output enhancer operates to provide the modified audio stream including inserted audio segments which were not received from the audio streaming server.

25 **[0016]** Still further in accordance with a preferred embodiment of the present invention the inserted audio segments include silence, pre-recorded audio segments and/or advertisements.

[0017] Typically the client includes a telephone. Additionally or alternatively the client includes a telephone and an IVR.

30 **[0018]** Additionally in accordance with a preferred embodiment of the present invention the client provides a real time output.

[0019] There is further provided in accordance with another preferred embodiment of the present invention a system for providing sophisticated seeking in an

audio stream. The system includes an audio streaming server providing an audio stream, an audio sampler, intermittently sampling portions of the audio stream, an audio sampling store, storing the portions sampled by the audio sampler and an audio stream portion navigating seeker operative to sequentially render the portions.

5 [0020] There is also provided in accordance with yet another preferred embodiment of the present invention a method for providing sophisticated seeking in an audio stream. The method includes the following steps: providing an audio stream, intermittently sampling portions of the audio stream, storing the intermittently sampled portions of the audio stream and seeking by sequentially rendering the portions.

10 [0021] Further in accordance with a preferred embodiment of the present invention the audio sampler is operative to sample complete phrases.

[0022] Still further in accordance with a preferred embodiment of the present invention the audio stream portion navigating seeker operates to insert at least one audible tone among the portions rendered thereby.

15 [0023] Further in accordance with a preferred embodiment of the present invention the audio sampler is voice command responsive.

[0024] Additionally in accordance with a preferred embodiment of the present invention the audio sampler samples portions which are selected generally periodically.

20 [0025] Typically the audio stream portion navigating seeker operates to render via a telephone. Additionally or alternatively, the audio stream portion navigating seeker is operative to render via an IVR and a telephone.

[0026] Further in accordance with a preferred embodiment of the present invention the system also includes a user operative, seeking responsive audio stream renderer, operating to render the audio stream beginning from a sampled portion selected by a user.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The present invention will be more fully understood and appreciated from the following detailed description, taken in conjunction with the drawings, in which:

30 [0028] Fig. 1 is a simplified block diagram of a system for audio listening in accordance with a preferred embodiment of the present invention;

[0029] Fig. 2 is a simplified block diagram of a process involved in enhanced buffering of audio in accordance with a preferred embodiment of the present invention;

[0030] Fig. 3 is a simplified flowchart illustration describing an example of a method of operation of the system shown in Fig. 2;

5 [0031] Fig. 4 is a block diagram illustration of a system and methodology for sampling audio phrases constructed and operative in accordance with a preferred embodiment of the present invention; and

[0032] Fig. 5 is a simplified flowchart illustration describing an example of a method of operation of the system shown in Fig. 4.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

10 [0033] The present invention provides a system and methodology for enhancing the streaming of audio and for sophisticated seeking within audio by sampling phrases from a given audio segment. The streaming of audio is provided by a streaming audio server, which may be any suitable device that provides a suitable audio output. Such
15 output may be live but need not necessarily be live. Examples of a suitable streaming audio server are an HTTP server, a disc cache and a REALAUDIO ® streaming server. A user's experience of audio can be enhanced by enabling the user increased control over navigation within an audio sample. To assist in understanding the present
20 invention, three simplified example sessions are described herein-below:

EXAMPLE SESSION 1

[0034] A telephone user who wishes to listen to audio may place a telephone call to the system of the present invention. The system of the present invention may send the streaming audio to the telephone. A user listening to the streaming audio may
25 experience periods of buffering due to network congestion.

[0035] The system of the present invention may monitor the rate of playback (bytes per second) and may detect the number of bytes that are needed in order to replenish the supply of received bytes in a buffer. Additionally, the system of the present invention may analyze the streaming audio and judge how often there will be a
30 chance to insert bytes within the streaming audio. The system of the present invention may also detect the best place to insert extra bytes and whether to insert background

noise, silence or third party background music. The system may then insert the extra bytes.

EXAMPLE SESSION 2

[0036] A user may wish to watch a one-hour video lecture on 'general relativity' via a telephone or any other suitable audio-visual client. The visual cues from the video may be minimal as the visual appearance of the lecturer and of the background may remain relatively unchanged for the entire duration of the lecture. The user may want to fast forward the lecture to the section on 'gravitation' that takes place approximately 20 minutes after the lecture starts. In this scenario, the present invention may provide a system and methodology for sophisticated seeking, preferably by sampling phrases from the audio, such as one phrase out of every seven phrases.

[0037] When seeking within the aforementioned audio segment, the user may be able to rewind or fast forward through numerous phrases of the lecture and hear only a sub-set of phrases. This sub-set of phrases may provide the user with enough audio cues and enough information to ascertain his "location" within the video, without the user having to listen to the entire audio segment.

EXAMPLE SESSION 3

[0038] A telephone user may want to access information such as news. The user may specify which news report he wishes to hear and may seek within the news report to find a specific segment. The user may employ the keypad and/or voice commands to navigate through the audio sample and preferably hears a sub-set of audio phrases that may enable the user to easily locate the specific segment.

[0039] Reference is now made to Fig. 1, which is a simplified block diagram of a system for enhanced user experience of audio listening. A user, employing a Telephone or other communicator 100, may place a telephone call to an Interactive Voice Response Unit (IVR) 110, which may contain a Buffer Module 120. The user may wish to access audio content on an audio streaming server 130, such as a Content Streaming Server, as described in US Patent Application No. 09/798,377 entitled "Telephone and Wireless Access to Computer Network-Based Audio".

[0040] The audio content may be streamed through a Temporal Modulator 140 and buffered with the Buffer Module 120. Should the user elect to seek within the audio stream, the Temporal Modulator 140 may be employed as described herein-below with

reference to Fig. 4. Furthermore, should network congestion require the system to re-buffer, the Buffer Module 120 may be utilized as described hereinbelow with reference to Fig. 2.

[0041] Reference is now made to Fig. 2, which is a schematic diagram of the functionality involved in buffering audio over the telephone and to Fig. 3, which is a simplified flowchart illustration describing an example of a method of operation of the system shown in Figure 2. In a preferred embodiment of the present invention, a user may wish to hear audio via the Telephone or other communicator 100. The user may connect via the Telephone or other communicator 100 (Fig. 1) to an IVR 110 (Fig. 1), using a Telephonic Interface 200, forming part of the IVR 110 and employ the IVR 110 for selecting specific audio content.

[0042] The user may select the audio content in various ways. For example, the user may have previously defined a set of personal preferences, which are stored on an easily accessible database (not shown) for later access via the telephone. Reference is made in this connection to assignee's copending U.S. Patent Application Serial No. 09/798,377 entitled 'Telephone And Wireless Access To Computer Network-Based Audio', the disclosure of which is hereby incorporated by reference.

[0043] The user may access the previously defined personal preferences through the IVR 110. The personal preferences may include Uniform Resource Locators (URLs) indicating the location of specific audio content on the Internet.

[0044] Another method for the user to select the content is to navigate menus of predefined audio content options, as illustrated in Fig. 3. The user may employ Dual Tone Multi Frequency (DTMF) and/or voice input to select the specific audio content [step 300]. Once the user has made a selection, the IVR 110 may retrieve the requested URL, referencing the specific audio content from a database (not shown) [step 310]. Next, the IVR 110 connects via the Internet to a Streaming Server 210 and receives an audio stream of the selected content.

[0045] Alternatively, IVR 110 may be preset to play the content.

[0046] The IVR 110 then requests the audio content from the Streaming Server 210 [Step 320]. Preferably, the Streaming Server 210 transmits a buffer of 8,000 bytes per second to the IVR 110 [step 330]. These 8,000 bytes correspond to roughly one second of audio. The telephone user 100 may be listening to this one second of audio

while new audio information may be transmitted and relayed to the Buffer Module 120. The Buffer Module 120 may contain an Incoming Buffer 220, a Rate Analyzer 230, an Optimal Detector 240 and an Optimal Inserter 250. The new information may be placed in the Incoming Buffer 220, where it may be copied to the Play Buffer 260 [step 340].

5 The Incoming Buffer 220 may have a fixed capacity for bytes whereas the capacity of the Play Buffer 260 typically is variable.

[0047] Should the new incoming information be delayed or slowed down, Rate Analyzer 230, which monitors the rate of playback (bytes per second), may detect that bytes are needed in order to replenish the supply in the play buffer. Optimal Detector 240, located within the IVR, which preferably analyzes the stream, may determine how often to insert these bytes within the present stream. The Optimal Detector 240 may detect background sound and locate the best place to insert these extra bytes. The Optimal Inserter 250 may determine whether to insert background noise, silence or third party background music in the form of a packet [step 350]. The packet may then be copied to the Play Buffer 260 [step 360].

[0048] Reference is now made to Fig. 4, which is a block diagram illustration of preferred functionality of the method and system of the present invention as shown in Fig. 1, and to Fig. 5, which is a simplified flow chart illustrating preferred operation of the system of Fig. 4, in accordance with a preferred embodiment of the present invention. In one embodiment of the present invention, a User, preferably employing a Telephone or other communicator 100, may want to navigate a sample of audio and receive audible cues enabling the user to better orientate himself.

[0049] As illustrated in Fig. 5, the user may first establish personal preferences, for instance, to play one phrase in every eight phrases of the audio sample [step 500]. Next, the Optimal Detector 240 may detect the beginning of a current phrase within the audio sample [step 510]. The Optimal Detector 240 may then seek within the audio sample and locate the end of the phrase [step 520]. An Optimal Extractor 400 may extract a copy of the phrase from the audio sample [step 530]. An Optimal Inserter 250 may insert an audible tone after the phrase [step 540]. After the audible tone is inserted, the audio sample and the audible tone may be transmitted to the user via Transmitter 410 [step 550].

5 [0050] The user may hear an audio segment including a sample phrase followed by an audible tone. While the user hears the audio segment, an Optimal Skipper 420 may seek forward or backwards within the audio sample and extract a further phrase [step 560]. The further extracted phrase may be distant from the previous phrase, for example, eight phrases later in the audio sample. As before, an audio segment may be constructed, including the extracted phrase followed by an audible tone. The audio segments are continuously relayed to the user in this manner.

10 [0051] It is appreciated that one or more of the steps of any of the methods described herein may be omitted or carried out in a different order than that shown, without departing from the true spirit and scope of the invention.

15 [0052] While the present invention as disclosed herein may or may not have been described with reference to specific hardware or software, the present invention has been described in a manner sufficient to enable persons of ordinary skill in the art to readily adapt commercially available hardware and software as may be needed to reduce any of the embodiments of the present invention to practice without undue experimentation and using conventional techniques.

20 [0053] While the present invention has been described with reference to one or more specific embodiments, the description is intended to be illustrative of the invention as a whole and is not to be construed as limiting the invention to the embodiments shown. It is appreciated that various modifications may occur to those skilled in the art that, while not specifically shown herein, are nevertheless within the true spirit and scope of the invention.

25 [0054] The present invention is not limited by what has been particularly shown and described hereinabove. Rather the scope of the present invention includes both combinations and subcombinations of features described hereinabove as well as modifications thereof which would occur to a person of skill in the art upon reading the foregoing description and which are not in the prior art.